their usefulness. For this reason, we investigated the use of a low-cost, handheld NIR spectrometer (SCiO, ConsumerPhysics) for field-based DMC prediction in cassava roots. Pilot investigation into the predictive effects of preprocessing techniques, number of roots sampled per plot, and within-root sampling location were used to develop a scanning and sampling protocol. Following this protocol, oven-dried measurements of DMC were paired with scans of roots of diverse clones from IITA (Nigeria), NaCRRI (Uganda), and Embrapa (Brazil) and grouped into training and test sets based on prediction scenarios common to plant breeding programs.Partial least squares regression models were evaluated for predictive ability, which ranged from R^2=0.51-0.85 depending on the cross-validation scheme. With appropriate calibration, this spectrometer will allow for field-based collection of spectral data with a smartphone for accurate DMC prediction, a step that could be easily integrated into existing harvesting workflows of cassava breeding programs. These and other NIRS models will be hosted on BreedBase to facilitate further analysis and incorporated into the PhenoApps suite of Android applications for plant phenotyping.

W290: Connecting Crop Phenotype and Genotype Data

Prioritization of Genetic Variants by Biological and Evolutionary Annotation: Functional Assessments in Diverse Maize Populations

Guillaume Ramstein[1], Sarah Jensen[2], Hai Wang[3], M. Cinta Romay[1] and Edward S. Buckler[1], (1)Institute for Genomic Diversity, Cornell University, Ithaca, NY, (2)Cornell University, Ithaca, NY, (3)Biotechnology Research Institute, Chinese Academy of Agricultural Sciences, Ithaca, NY

Genomic prediction is a useful tool for estimating genetic values and quickly selecting individuals within breeding programs. However, genomic prediction models typically identify effects of genomic markers that are confounded by genetic linkage to the actual causal variants. Therefore, such models typically cannot generate useful insight about the genetic architecture of traits, and tend to lose accuracy across population backgrounds (e.g., prediction in a bi-parental breeding population based on a model trained in a different diverse population). Based on two panels of 1,106 and 1,640 maize hybrids, we show the usefulness of annotations about gene action (additive and dominance effects), genic regions (distance to genes), and evolutionary constraint (cross-species allelic conservation) for genomic prediction across population backgrounds, whereas annotations about marker effects from genome-wide association studies could not contribute to any increase in prediction accuracy. We further show the limitations of current methods for incorporating annotation information in genomic prediction models, due to low density of marker assays, and limited statistical power for estimating the importance of various annotations. To address these limitations, we present an innovative imputation method based on haplotype graphs (the Practical Haplotype Graph), which should provide an exhaustive assay of genetic variability around genes. Furthermore, we present machine learning methods for incorporating various biological annotations of genetic variants (recombination rate, chromatin openness, expression effects, gene ontology, etc.) based on evolutionary constraint. Our results show promising avenues for effectively incorporating biological information into genomic models and increasing their robustness in cross-population prediction.

W291: Connecting Crop Phenotype and Genotype Data

Empowering Global Rice Breeding Programs using Genomic Selection

Parthiban Prakash[1], Juan David Arbelaez Velez[2], Holden Verdeprado[3], Vitaliano Lopena[4], Rose Imee Zhella[4], Jerome Bartholome[5], Jessica Rutkoski[6], Rosemary Murori[7], Alexis Ndayiragije[8], Mohammad Rafiqul[9], Sanjay Kumar Katiyar[10] and Joshua N. Cobb[11], (1)International Rice Research Institute, Los Banos, Philippines, (2)University of Illinois Urbana Champaign, (3)International Rice Research Institute, Los Baños, Philippines, (4)International Rice Research Institute, Philippines, (5)International Rice Research Institute | CIRAD, Philippines, (6)University of Illinois Urbana Champaign, Los Baños, (7)International Rice Research Institute, Tanzania, United Republic of, (8)International Rice Research Institute, Mozambique, (9)International Rice Research Institute, Bangladesh, (10)International Rice

Research Institute (IRRI), Manila, India, (11)International Rice Research Institute, Los Baños, Laguna, Philippines

Rice is the staple food for half of the global population and irrigated rice contributes 70% of total rice produced. Given the strategic importance of this market segment to global food security, the irrigated rice breeding program at IRRI is mandated to breed varietal rice and empower rice breeding programs in South Asia and Eastern and Southern Africa. As public sector breeding budgets are insufficient to adequately test all new lines across such a varied environmental landscape, advanced molecular tools such as genomic selection (GS) are critical to achieving high levels of selection intensity and selection accuracy within each region. For the past two years, the irrigated lowland breeding program at IRRI has distributed carefully selected 'estimation sets' containing a few hundred breeding lines to global partners in Africa and South Asia with the intention of using genomic prediction to evaluate several thousand selection candidates that have only been genotyped. Combining different software tools such as B4R (phenotypic data management), GOBii (genotypic data management) and ASReml-R (modeling) into an analytical workflow, we have generated initial prediction accuracies for grain yield, plant height, and flowering time which are routinely used in the program to make optimized breeding decisions for each of these disparate global environments. An overview of the results will be presented as well as a discussion about the future integration of sparse testing designs, multi-environmental models, use of weather data through machine learning, and phenotyping with unmanned aerial vehicles (UAV).

W292: Connecting Crop Phenotype and Genotype Data

Utilizing Phenomics, Genomics, and Deep Learning Methods to Accelerate the Development of Climate Resilient Crop Varieties

José Crossa, International Maize and Wheat Improvement Center (CIMMYT), Mexico City, Mexico, Osval Montesinos-Lopez, University of Colima, Colima, CL, Mexico, Abelardo Montesinos-Lopez, University of Guadalajara, Guadalajara, JA, Mexico, Jaime Cuevas, University of Quintana-Roo, Chetumal, QR, Mexico, Paulino Perez-Rodriguez, Colegio de Postgraduado, Texcoco, EM, Mexico, Margaret Karuse, Researcher, Texcoco, EM, Mexico, Juan Burgueño, CIMMYT, El Batan, Mexico and Fernando Toledo, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Hyperspectral reflectance phenotyping and genomic selection are two technologies that have the potential to increase prediction accuracy for grain yield under different environmental conditions. Hyperspectral cameras quantify canopy reflectance associated with numerous biophysical and biochemical processes in plants, and genomic selection models utilize dense molecular markers or/and pedigree information to predict the breeding values of lines. Authors have proposed Bayesian functional regression models that take into account all available bands, genomic or pedigree information, the main effects of lines and environments × environment interaction (G×E) and wavelength × environment interaction (B×E). Results show that models with B×E interaction terms were the most accurate models, whereas the functional regression models (with B-splines and Fourier basis) and the conventional models performed similarly in terms of prediction accuracy. However, the functional regression models are more parsimonious and computationally more efficient because the number of beta coefficients to be estimated are lower (number of basis) than estimating all regression coefficients for all bands. Also other studies have proposed hyperspectral reflectance-derived relationship matrices to model the genotype × environment (G×E) interactions across environments. Multi-kernel models combining marker/pedigree information with hyperspectral reflectance phenotypes had the highest prediction accuracies. Last years several genomic-enabled predictions models have been developed with increasing prediction accuracy over the genomic best linear unbiased estimator (GBLUP). These methods are the Gaussian kernel (GK) and Arc-cosine (AK) kernel. The AK kernel method emulates the deep machine learners (DL) and are much easier to implement and with less computational time and similar or slightly higher genomic-enabled prediction accuracy than the GK kernel. Genomic DL methods have been useful for assessing big data comprising multi-trait multi-environment. Implementing the multi-trait DL models is feasible but challenging due to the large number of hyper-parameters involved.

W293: Connecting Crop Phenotype and Genotype Data